

---

## **Electronic Enforcement Screening Pilot Project**

---

**February, 2000**

**U.S. Environmental Protection Agency  
Office of Enforcement and Compliance Assurance  
Office of Planning and Policy Analysis**

## ACKNOWLEDGMENTS

The principal author of this report is Nicholas C. Franco of the Office of Planning and Policy Analysis in the U.S. EPA's Office of Enforcement and Compliance Assurance. The Electronic Enforcement Screening Pilot Project was a jointly funded effort of the Toxic and Pesticides Enforcement Division, U.S. EPA Office of Regulatory Enforcement, and the Office of Planning and Policy Analysis. Appreciable assistance was received from Seth Heminway, Manufacturing, Energy, and Transportation Division, U.S. EPA Office of Compliance; and the EPCRA § 313 Enforcement Coordinators in Regions I and IV. Abt Associates Inc., Cambridge, Massachusetts, compiled and analyzed data, formatted sector and facility specific deliverables, and authored technical reports (see appendices) under Contract #68-W98-005, WA 1-07 and WA-2-07.

For additional information or questions concerning this report, please contact Mr. Franco via telephone at (202) 564-0113 or Email at *franco.nicholas@epa.gov*.

## TABLE OF CONTENTS

EXECUTIVE SUMMARY .....	iii
INTRODUCTION .....	1
PROJECT DESCRIPTION .....	2
PROJECT EXECUTION .....	4
Choosing a Sector to Investigate .....	4
Universe Definition .....	5
Data Collection .....	7
Data Analysis .....	12
PROJECT EVALUATION .....	17
Region 4 Review of the Wood Preserving Sector .....	17
Region 1 Review of the Paper & Paperboard Sector .....	18
LESSONS LEARNED .....	20
RECOMMENDATIONS .....	24
APPENDIX A	
INFORMATION SOURCES REVIEWED FOR THE PILOT PROJECT .....	26
APPENDIX B	
FACILITY RANKING TABLES	
Appendix C	
Region I Report on Review of Paper and Paperboard Sector Targets	
APPENDIX D	
ABT MEMORANDUM, “TOXIC RELEASE INVENTORY (TRI) SECTOR ANALYSIS	
APPENDIX E	
ABT MEMORANDUM, “ELECTRONIC ENFORCEMENT PROJECT: WOOD PRESERVING SECTOR	
FACILITY REPORTS PACKAGE	
APPENDIX F	
ABT MEMORANDUM, “ELECTRONIC ENFORCEMENT PROJECT: PAPER SECTOR FACILITY	
REPORTS PACKAGE”	
APPENDIX G	
ABT MEMORANDUM, “ELECTRONIC ENFORCEMENT PROJECT: CANNED AND FROZEN	
POULTRY SECTOR FACILITY REPORTS PACKAGE”	

## Executive Summary

The impetus for the Electronic Enforcement Screening Pilot Project grew out of an offshoot of the Sector Facility Indexing Project (SFIP), which attempted to leverage inconsistencies in data collected for the project to identify Toxic Release Inventory (TRI) non-reporters. The ad hoc methodology for identifying TRI non-reporters was marginally successful. In the fourth quarter of 1997 the Office of Planning and Policy Analysis, and the Toxics and Pesticides Enforcement Division within the Office of Regulatory Enforcement sponsored the Electronic Enforcement Screening Pilot Project to determine whether a more deliberate and focused effort at identifying potential non-reporters, via the aggregation and analysis of electronic data, would be more successful. The primary goal of the project was to define the universe of facilities in a sector as completely as possible, and to generate an up-to-date and accurate list of potential TRI non-reporters in that sector. A secondary goal was to devise a general methodology, and identify a set of data sources, that could be used to implement future electronic enforcement projects.

### Project Description

**Universe Definition:** The first step in targeting non-reporters was to compile a comprehensive list of facilities in the universe, not just those in EPA databases, defined by the SIC code under investigation. To ensure the comprehensiveness of the list, information from EPA and outside databases was combined.

**Data Collection:** Information collection was roughly divided into two phases, the first phase more exploratory, cataloging information sources on a particular sector and putting together a profile of the sector (classes of products produced, major chemicals used, etc). The second phase focused on gathering direct information on whether facilities in the sector were over the TRI reporting thresholds for number of employees and annual chemical usage.

**Data Analysis:** Data analysis focused primarily on finding methods of estimating whether a facility had exceeded the TRI chemical usage threshold. For one sector existing models were used to estimate facility chemical usage. For the other sectors statistical techniques were used to give an indication of whether a facility had exceeded the chemical usage threshold.

### Lessons Learned

**Choosing a Sector to Target is not Straightforward:** choosing an appropriate sector for electronically-based screening and targeting can be difficult and resource intensive. It is not advisable to choose a sector without first gathering general background information on the sector, estimating the magnitude of the potential non-reporting problem, characterizing the sector in terms of chemical usage, and determining the availability of quality information sources.

**Not All Sectors As Defined by SIC Code Lend Themselves to This Type of Analysis:** Sectors that are homogeneous with respect to chemicals used and production processes employed are best suited to the electronic screening process. In many cases, sectors described by four-digit SIC codes are too diverse for them to be good candidates. Additionally, sectors that use listed chemicals in

the production process or integrate them into the final product are easier to model than those sectors that use listed chemicals in a secondary operation.

**Universe Definition is Labor Intensive:** The current state of data sources external to the EPA does not support a straightforward and relatively automatic compilation of facilities in a sector. The utility of commercial and industry-specific databases is limited by data quality problems. Specifically, the inaccurate assignment of SIC codes to facilities by commercial databases. The universe definition process was made more difficult because the integration of information from EPA and outside data sources was largely a manual process.

**Information in Commercial Databases Does Not Support Facility-level Analysis:** Commercial databases rarely contain facility-level information required by the electronic screening process, and much of the information they do contain is estimated based on industry averages (i.e., may not be a reliable indicator of facility-specific sales or production volume).

**Other Federal Agencies Were Not Useful Information Sources:** With the exception of some EPA databases, and Census Bureau numbers on the size of industry sectors, federal agency sources were of little utility because they lacked sufficient specificity, or were not comprehensive.

## INTRODUCTION

The impetus for the Electronic Enforcement Screening Project grew out of similar work that was done as an offshoot of the Sector Facility Indexing Project (SFIP) in 1995 and 1996. SFIP integrated information on permits, enforcement and compliance history, environmental releases, and general background information for facilities in five industrial sectors. When pulling together information from different data sources to create facility profiles, inconsistencies were noticed that led people to believe that some facilities were potential EPCRA § 313, Toxic Chemical Release Inventory (TRI) non-reporters. For example, in the petroleum refining sector a number of facilities had reported to the Department of Energy's Petroleum Supply Reporting System, but had not filed a TRI Form R. Given the congruence between reporting to the two programs and the reported production capacity, a number of facilities were identified as potential non-reporters. Similar inconsistencies between production statistics, reporting to media programs, and TRI reporting were used to identify potential non-reporters in other SFIP industrial sectors.

The Office of Planning and Policy Analysis (OPPA) and the Toxics and Pesticides Enforcement Division (TPED) in the Office of Regulatory Enforcement (ORE) worked together to distribute the lists of potential TRI non-reporters to Regional EPCRA § 313 Enforcement Coordinators for investigation. The ad hoc methodology for identifying potential TRI non-reporters was marginally successful. Many of the facilities were later found to have reported to TRI, did not meet one of the TRI reporting thresholds, were out of business, or had not yet begun production.

The Electronic Enforcement Screening Pilot Project was initiated in response to some of the problems encountered in this ad hoc approach to identifying potential non-reporters as part of the SFIP. The overarching goal was to determine whether a more deliberate and focused effort at identifying potential TRI non-reporters could overcome some the problems encountered with the potential non-reporters identified as an offshoot of the SFIP.

### Summary of TRI Reporting Requirements

Established under Section 313 of the Emergency Planning and Community Right-to-Know Act of 1986 (EPCRA), the Toxic Chemical Release Inventory (TRI) is a publicly available database that contains specific toxic chemical release and transfer information from manufacturing facilities. A facility must report to TRI if it:

- C Conducts manufacturing operations within Standard Industrial Classification (SIC) codes 2000 - 3999,
- C Has 10 or more full-time equivalent employees, and
- C Manufactures (defined to include importing) or processes more than 25,000 pounds; or otherwise uses more than 10,000 pounds of any TRI-listed chemical during the calendar year.

As of November 30, 1994, there were 643 TRI-listed chemicals and chemical categories. According to the *Toxic Chemical Release Inventory Reporting Form R and Instructions* (1996), the term "manufacture" means to produce, prepare, compound, or import a listed toxic chemical. The term "process" means the preparation of a listed toxic chemical, after its manufacture, for distribution in commerce. The term "otherwise use" encompasses any activity involving a listed toxic chemical at a facility that does not fall under the definitions of "manufacture" or "process."

Additional information on specific examples and exemptions can be found in the *Toxic Chemical Release Inventory Reporting Form R and Instructions*.

## PROJECT DESCRIPTION

The Electronic Enforcement Screening Project was initiated in the fourth quarter of 1997, jointly funded by OPPA and TPED. The original plan was to screen four industrial sectors. OPPA and TPED chose the first two sectors to be investigated, wood preserving, defined by Standard Industrial Classification (SIC) code 2491; and paper and paper board production, SIC codes 2621 and 2631 respectively. In December of 1997 a memorandum from ORE and OPPA was sent to the regional EPCRA § 313 Enforcement Coordinators announcing the project and asking them to nominate two more sectors for screening. Based on regional suggestions the canned and frozen poultry sector (SIC 2015) was the third sector investigated.<sup>1</sup> Time and budget considerations did not allow a complete analysis of the third sector, and prohibited investigation of a fourth sector.

Though the overarching goal of the project was to attempt to improve upon past attempts at identifying potential non-reporters via the aggregation and analysis of electronic data, it can be further broken down into primary and secondary operational goals. The primary goal of the project was to define the universe of facilities in a sector as completely as possible, and to generate an up-to-date and accurate list of potential TRI non-reporters in that sector. A secondary goal was to devise a general methodology, and identify a set of data sources, that could be used to implement future electronic enforcement projects. The three major steps in sector investigations -- universe definition, data gathering, and data analysis -- are described in more detail below.

**Universe Definition:** The first step in targeting non-reporters was to compile a comprehensive list of facilities in the universe defined by the SIC code under investigation. Thoroughly defining the universe is important to ensure that the scope of the project includes more than just the facilities that are currently in EPA databases. To ensure the comprehensiveness of the list, information from EPA and outside databases was combined. The list was started with facilities that had reported to the TRI Database (TRIS) in the most recent year for which data was available, and several years prior. Information was also included on facilities in other EPA databases that reported the SIC code under investigation. Outside information sources generally fell into two broad categories, commercial databases such as Dun & Bradstreet (D&B), and industry specific information sources such as a survey of wood preservers by the American Wood Preserving Institute (AWPI), or a listing of facilities in the pulp, paper and paperboard industries from the Lockwood-Post Directory of Pulp, Paper, and Allied Trades. Once all the information sources were pulled together, duplicate facilities were removed, and an attempt was made to link each facility via the Facility Indexing System (FINDS) to facilitate further data gathering through OECA's Integrated Data for Enforcement Analysis (IDEA) database.

---

<sup>1</sup> See Appendix G for Abt Associates, Inc. memorandum on the analysis of the canned and frozen poultry sector.

**Data Collection:** Information collection was roughly divided into two phases, the first phase more exploratory, cataloging information sources on a particular sector and putting together a profile of the sector (classes of products produced, major chemicals used, etc). EPA and industry information sources were also searched for studies, guidance documents, and regulatory impact analyses which would be helpful in determining whether a facility used specified chemicals above TRI reporting thresholds. The primary goal of the first phase was to inform and narrow the scope of information collection activities in the second phase.

The second phase focused on gathering direct information on whether facilities in the sector were over the TRI employee and annual chemical usage thresholds. Information on number of employees was readily available from commercial databases such as D&B. Direct information on chemical usage at the facilities was not available for any of the sectors investigated. For the wood preserving sector an EPA regulatory impact analysis was used to estimate chemical usage at facilities. For the other sectors, information was gathered on production volume, sales, and other readily available company and facility statistics that might be correlated with chemical usage. The following section on data analysis explains in more detail how this information was used to estimate facility chemical usage.

**Data Analysis:** Data analysis focused primarily on finding methods of estimating whether a facility had exceeded the TRI chemical usage threshold. Models developed as part of a regulatory impact analysis for the listing of wood preserving wastes were used to estimate annual chemical use for facilities in the wood preserving sector.<sup>2</sup> An estimate of annual chemical usage was not possible for the other sectors; therefore, in those cases statistical methods were relied upon. For example, average sales and production statistics were computed for reporting facilities. If the values for potential non-reporters were above the average then it may be an indication that the facility should be reporting. More sophisticated statistical techniques were also used to determine whether there was a relationship between a known variable (such as sales, production volume, or number of employees) and annual chemical usage for reporting facilities. This could then be used to estimate chemical usage for non-reporters. No useful relationship between a known variable and chemical usage was found.

---

<sup>2</sup>U.S. EPA, Office of Solid Waste. *Regulatory Impact Analysis for the Final Listing of Certain Wood Preserving Wastes* ("RIA"), prepared by ICF Incorporated for the Economic Analysis Staff, Office of Solid Waste, U.S. EPA, November 1990.



## **PROJECT EXECUTION**

This section of the report evaluates the success of the pilot project in carrying out the three major process steps (universe definition, data collection, and data analysis) outlined in the previous section. The methods used to carry out these steps are summarized, and difficulties and analytical limitations are also identified. The next section, Project Evaluation, evaluates how successfully the project identified TRI non-reporters.<sup>3</sup>

Overall, the process of generating a list of potential TRI non-reporters was much more time and resource intensive than originally anticipated. Researching potential sectors in order to find suitable candidates for the pilot was an unanticipated drain on resources, and comprised an additional step prior to universe definition. Additionally, universe definition was a manual and time consuming process. These two factors were primarily responsible for the decision to scale back the project from four sectors as originally planned, to two sectors, with a third only partially completed.

### **Choosing a Sector to Investigate**

The first two sectors were chosen by headquarters staff because facilities in these two sectors were known to use large quantities of TRI listed chemicals, and were deemed ripe for analysis. Data from the 1992 Census of Manufacturers reported 486 wood preserving facilities, with approximately half of exceeding the TRI reporting threshold of 10 or more full-time employees. Coupled with the knowledge that wood preserving is a chemical-intensive industry (with the chemicals predominantly used in the sector being TRI reportable), wood preserving was chosen as the first sector to investigate as part of the pilot project. The paper and paperboard sector was chosen because initial analysis of the sector as part of SFIP had shown it to be promising for further investigation. SFIP had identified 300-400 papermaking facilities in the United States. Of these, approximately 100 had not reported to TRI in the past (i.e., there was a large population of potential non-reporters). Initial analysis of the sector showed that a variety of TRI reportable chemicals were used during paper production. Additionally, there was a potential to save money by using the work already done by the contractor to catalog facilities in the papermaking sector for SFIP. Little or no attempt was made to determine the availability of information sources for either of these sectors before they were selected.

Good information sources were available for the wood preserving sector. The Office of Compliance had recently compiled a list of facilities in the wood preserving sector, economic models of the sector were available in a regulatory impact analysis, and a survey of the sector (including information on chemical type used and production capacity) was available from the American Wood Preserving Institute. Early on in the process of gathering information for the second sector, paper and paperboard production, it became apparent that sources of information of similar quality were not available. The inability to find information sources for the second sector was taken into account when

---

<sup>3</sup> See Appendix D for a more in-depth, technical description of the methodology used in the electronic enforcement screening pilot project.

choosing the remaining sectors to be investigated. The availability of quality information on a sector was added as a criteria for choosing the remaining sectors.

The remaining two sectors were chosen from a pool of suggestions by the EPCRA §313 Regional Enforcement Coordinators. A number of suggested sectors were rejected because of potential conflict with existing EPA initiatives (the electroplating sector was rejected because an Agency-wide Common Sense Initiative already existed in this sector; likewise an investigation of the inorganic chemical sector was rejected because the Office of Compliance was already conducting an investigation of this sector). Several regions nominated some portion of the food processing industry. The food processing industry was too large to investigate in its entirety for the purposes of this pilot, so a sector or sectors of reasonable size had to be identified. Some of the criteria used for evaluating sectors were: past TRI reporting record, availability of information sources on the sector, whether the sector consistently employed a process and/or chemicals which would require filing a TRI report if reporting thresholds were exceeded, and number of facilities in the sector.

Extensive analysis was conducted on the food processing industry in order to identify an appropriate sector for analysis. The food processing industry, SIC major group 20, is divided into nine industry groups, with 48 four-digit SIC code classifications. Census data and a compliance report on the food processing industry done by the Office of Compliance in April, 1997 was used to narrow the field of investigation to three of the nine industry groups. The three groups: meat products, grain and mill products, and sugar and confectionery products were chosen because they were of manageable size, had average number of employees at or above the TRI reporting threshold, and did not have a high TRI reporting rate (i.e. potentially may have a large number of TRI non-reporters). The canned and frozen poultry sector, SIC 2015, was chosen at the conclusion of this analysis.

Initial analysis of a suitable candidate for a fourth sector was conducted. However, due to the unexpectedly extensive analysis required to identify the canned and frozen poultry sector, and finding information sources for both the second and third sectors, it became clear that there would not be sufficient resources left to conduct a complete analysis of a fourth sector.

## **Universe Definition**

More time and effort was spent in attempting to comprehensively define the universe of facilities for the first sector, wood preserving, than was for the other sectors. Lessons learned in this effort led to scaling back universe definition efforts for the other sectors.

The universe of facilities was generally described by the Standard Industrial Classification (SIC) code definition corresponding to the industry under investigation. The plan for compiling a comprehensive list of facilities that fit the SIC code description was to bring together numerous and disparate sources of facility information (EPA, commercial, and industry databases) to ensure the list was comprehensive.

For the wood preserving sector, SIC 2491, a list of facilities from an industry survey conducted by the American Wood Preserving Institute (AWPI) formed the nucleus of the facility list. Lists of facilities

with the corresponding SIC code were pulled from the TRI database (TRIS), RCRA Biennial Reporting System (BRS), CAA database (AIRS), and from the CWA database (PCS). Facility lists from TRIS and BRS were used, while information from other EPA databases was not included. This nucleus was augmented with facilities from the commercial database Dunn & Bradstreet (D&B).

Merging lists of facilities from EPA, commercial, and industry databases proved to be very time and labor intensive. Removal of duplicate facilities from the combined lists was a manual process due to inconsistencies in the presentation of name and address information. In many cases address information, or some other data field, was not available for a facility in D&B. In those instances the data was filled in by manually searching for the information in the Manufacturers News database. Also, for the wood preserving sector, D&B showed several hundred more facilities in the sector than the AWPI survey, or an estimate of the size of the wood preserving sector generated for a 1996 RCRA compliance guide for the wood preserving sector.<sup>4</sup> This raised the concern that numerous facilities were incorrectly identified as wood preservers by D&B.

Industry sources of information found for the other sectors investigated as part of this project were not as comprehensive as the AWPI survey. Most often industry sources only include lists of members of a specific industry organization, not a comprehensive list of all facility locations. If descriptive information on the sector was included in an industry publication most often it was aggregate information or sector averages (e.g. total sector production, average sales, average number of employees), and not facility-specific (e.g. production capacity and chemical types used) as in the AWPI survey.

Information collected by other federal regulatory agencies was also investigated to determine if it could be of use in defining the universe of facilities, or provide other useful descriptive information. Inquiries were made at the Occupational Safety and Health Administration (OSHA), the Department of Transportation (DOT), and the Food Safety Inspection Service at the Department of Agriculture. None of these agencies provided sources of information useful to this effort. Census data was used to help choose a sector to investigate in the food processing industry.

Given the time and effort required to merge lists of facilities from divergent sources, and the questionable nature of facility information in commercial databases, it was decided after the first sector to limit the number of information sources used to define the universe. The decision was made to look for, and focus on, a few high quality sources of information for the remaining sectors.

---

<sup>4</sup>U.S. EPA, Office of Enforcement and Compliance Assurance, Office of Compliance. *Wood Preserving Resource Conservation and Recovery Act Compliance Guide*, June 1996 (EPA-305-B-96-001)

## Data Collection

The primary focus of the data collection effort was to find industry-specific sources of information which would facilitate the process of determining whether a particular facility was over TRI reporting thresholds. The first phase of the data collection process involved cataloging and evaluating a wide variety of information sources, and putting together a sector profile (classes of products produced, major chemicals used, etc). Three types of information sources were searched: Federal government sources, commercial databases, and industry specific information sources (industry associations, trade groups, industry specific databases, etc.). Some of the criteria used to determine whether an information source would be used to support a sector analysis:

- **Type of information available:** did the source provide primary or secondary information on TRI reporting thresholds? Primary information would be pounds of chemical used or number of employees, secondary information would be sales, production volume, or some other variable which could be associated with threshold values.
- **Data Quality:** how was data collected, how often was it updated, how were SIC codes assigned, how much information was self-reported vs. estimated by another source
- **Data Gaps:** coverage of the universe under investigation.
- **Is the information company or facility specific?**
- **Ease with which information is searched and linked:** at a minimum did the source contain complete address information or some other way to uniquely identify facilities?
- **Data Format:** paper or electronic format?
- **Cost**

Information sources which characterized the industry sector in general, describing processes or chemicals which were typically used in the industry, were also investigated.

The table on following page lists the information sources evaluated for the three sectors screened as part of the pilot. Not all sources listed below were used in the project. Those that were used have a check in the far right column. For more detailed information on each of the data sources see Appendix A.

### *Information Sources Evaluated for the Electronic Enforcement Pilot Project*

<b>Category</b>	<b>Information Provider</b>	<b>Database / Information Source</b>	<b>Used</b>
<b><i>Federal Government</i></b>	Environmental Protection Agency	Toxic Chemical Release Inventory System (TRIS)	<b>U</b>
		Biennial Reporting System (BRS)	<b>U</b>
		Automatic Information Retrieval System (AIRS)	<b>U</b>
		Permit Compliance System (PCS)	<b>U</b>
		Regulatory Impact Analysis for the Final Listing of Certain Wood Preserving Wastes, 1990	<b>U</b>
		Integrated Data for Enforcement Analysis (IDEA)	<b>U</b>
	U.S. Census Bureau, Department of Commerce	Provided information on the size of industry sectors	<b>U</b>
	Occupational Health and Safety Administration	Inquired about lists of facilities in the sectors being investigated	
	Food Safety and Inspection Service, Department of Agriculture	Meat and Poultry Inspection Directory	
	Department of Transportation	Inquired about hazardous waste manifests	
<b><i>Commercial</i></b>	Dun & Bradstreet (D&B)	Dun Market Identifiers	<b>U</b>
	Manufacturer's News Inc.	Manufacturer's News	<b>U</b>
	American Business Information Marketing, Inc.	American Business Directory	
	Database Publishing Co.		
	Gale Research Inc.	Ward's Business Directory of U.S. Public and Private Companies	
<b><i>Industry Specific</i></b>	American Wood Preservers Institute (AWPI)	The 1995 Wood Preserving Industry Production Statistical Report	<b>U</b>
	Miller Freeman Publishing	1997 Directory of the Wood Products Industry	<b>U</b>
	Random Lengths Publishing	1997 Big Book	
	Southern Lumber Manufacturers Association	1997 Members Directory	
	Western Wood Preservers Institute	Sources of Supply of Pressure Treated Wood	
	Southern Forest Products Association	1997 Buyer's Guide	
	Lockwood-Post	Directory of Pulp, Paper, and Allied Trades, 1996	<b>U</b>

	Edward E. Judge & Sons, Inc.	The Directory of the Caning, Freezing, Preserving Industries	U
--	------------------------------	--	---

**Federal Government Sources:** With the exception of some EPA databases, federal government sources were of little utility because they either lacked sufficient specificity or were not comprehensive. The TRIS database was used to identify which facilities in the previously defined sector universe had reported, thus making the remaining facilities potential non-reporters. Information on reporters in a sector was also used to develop descriptive statistics on a typical reporter, which was used in conjunction with other information to give an indication of whether potential non-reporters should have reported. Other EPA databases such as BRS, PCS, and AIRS were used to augment the universe of facilities for a sector, pulling out those facilities in the database which had the same SIC code as the sector under investigation. An unsuccessful attempt was made to use information from other EPA databases to assist in the determination of whether a particular facility was over TRI reporting thresholds.<sup>5</sup> The utility of EPA databases was hampered by the inability to easily link facilities across EPA databases, necessitating significant effort to manually go through the data and individually link facilities. The lack of overlap between the information in EPA program databases and lists of potential non-reporters also limited their utility.

The Occupational Health and Safety Administration's (OSHA) Office of Statistics was approached to determine whether they had information which would help define the universe of facilities in a sector. OSHA does not maintain a comprehensive list of facilities in any sector. Instead, several priority sectors are chosen at random from the 200 "worst" industries based on health and safety compliance. Inspection targets are then chosen at random from Dun & Bradstreet, or based on complaints received.

The Department of Agriculture's Food Safety and Inspection Service (FSIS) annually compiles a meat and poultry inspection directory, which lists all the facilities inspected by FSIS. This should be a comprehensive source of facility information for this sector since an inspector is required to be on-site at each of the processing facilities. However, the directory does not contain SIC codes, or enough descriptive information about facility operations, to determine whether it was in the sector under investigation. Staff at FSIS were approached to determine if additional facility information was available that was not published in the annual directory that would enable the directory to be used as part of the pilot project, and were also asked if the information was available in an electronic format to facilitate using the information. FSIS staff were not forthcoming, and a Freedom of Information Act request did not yield sufficient additional information to enable use of the information in the inspection

---

<sup>5</sup> See Appendix F for a more complete discussion of the attempt to use other EPA databases to determine whether a facility in the paper/paperboard sector had exceeded the chemical use threshold.

directory.

**Commercial Sources:** Commercial databases provided three types of information: names and addresses of facilities to help define the universe under investigation, number of employees, and secondary information such as sales and production volume. However, most of the information in commercial databases is company specific, not facility specific, which limits its utility in projects of this type where facility specific information is required to identify potential non-reporters. A more significant problem was the inaccuracy of SIC code designation in commercial databases. SIC codes are assigned by the database publishers based on information collected about the company, but the incorrect SIC code is frequently assigned. This problem is exacerbated by the inability to distinguish between production facilities (which may be required to report under TRI) and other types of company operations (which likely do not have to report under TRI) both of which were given the same SIC code within the commercial database. For example, the AWPI survey identified 451 facilities in the wood preserving sector, while the D&B database listed 736 facilities in the sector. No other data source investigated listed more than 450 facilities in the sector. This has the potential corrupt the universe under investigation by adding facilities that have been assigned the wrong SIC, or that are not production facilities. This can lead to falsely identifying facilities as non-reporters, and corrupt the data pool on reporters and potential non-reporters making statistical methods for identifying non-reporters difficult.

The two commercial databases used for this pilot project were Dun & Bradstreet's Dun's Market Identifiers (DMI) and the Manufacturer's News database. With both databases there were problems with incomplete company records (missing sales, employee, address information, etc.). DMI served as the primary commercial database for the pilot project, and Manufacturer's News was employed in an attempt to fill in gaps in DMI. Linking facilities from DMI to EPA databases was a manual process, requiring matching of name and address information between databases.

Another potential pitfall of commercial databases is that much of the information they contain is not directly reported by the company, but instead is estimated by the database company. For example, as of January 1998, only 12% of the over 11 million records in DMI have company reported sales information, while 69% of the records have sales figures estimated by DMI, and the remaining 19% have no sales information at all.<sup>6</sup> Sales estimates are based on an industry per employee sales average. Thus, individual company estimates are dependent on SIC code designation, and the number of reported employees.

For a more complete discussion of the pros and cons of large commercial databases see the article "A Clash of Titans: Comparing America's most Comprehensive Business Directories," Database, June, 1998, <http://www.onlineinc.com/database/DB1998/lavin6.html>.

**Industry Specific Sources:** The industry specific information sources investigated as part of the

---

<sup>6</sup>A Clash of Titans: Comparing America's Most Comprehensive Business Directories, Database, June, 1998, <http://www.onlineinc.com/database/DB1998/lavin6.html>

pilot project can be broken into two categories: trade and industry association sources, and sector specific commercial sources. Typically, industry and trade association sources were narrow in scope, covering only a particular geographic region or members of the association in question. The most common types of information provided were name and address, and type of products produced. This was in keeping with the apparent mission of most trade and industry associations to promote member businesses, not to provide a comprehensive accounting of the industry. A few of these sources did provide summary statistics on the industry, but these were usually limited to average company sales or production numbers, types of process employed, or products produced in the industry. All of these sources were only available in paper format, which meant information had to be manually entered and linked to other information sources in order to be used. Given this, the spotty coverage, and lack of useful information, most industry and trade association information sources were of little value to the pilot project. The one exception was the American Wood Preserving Institute's 1995 Wood Preserving Industry production Statistical Report. The report surveyed the entire industry providing name and address, chemical process type used, and plant production capacity. The report also provided useful summary information on number of employees, average sales, and average shipments measured in dollars for the five years preceding the report.

Industry-specific commercial sources were found for the other two sectors screened. The Directory of Pulp, Paper, and Allied Trades published by Lockwood-Post was used as the primary information source for the paper/paperboard sectors. The directory listed name and address information, facility-specific production volume, production equipment, products, steam/power generation, water usage, and effluent treatment. Review of the paper/paperboard sector information by Region I showed that the industry-specific database did a much better job than the large commercial databases of assigning facilities the correct SIC code. However, a number of the facilities were found to have been out of business for some time, which is hard to explain for a database that is updated annually.

The Directory of the Canning, Freezing, Preserving Industries was used as the primary information source for the canned and frozen poultry sector, and was available in both paper and electronic formats. The directory was organized by company, but had separate entries for each production facility which listed name, address, products produced, number of employees, and SIC code. Company-level statistics on production volume are available (for canners and glass packers range of cases packed is designated, for frozen items range of pounds is designated), however this is not broken down by facility for companies that have multiple locations. Since analysis was not completed for this sector, no information is available to evaluate the quality of this data source.



## Data Analysis

A hierarchy of methodologies was employed to estimate whether a facility had exceeded TRI employee and chemical usage reporting thresholds,<sup>7</sup> and generate the list of potential EPCRA § 313 non-reporters. The methodologies are listed in order of preference below.

- **Collection of Primary Information** - for the pilot project primary information was defined as that which would enable a direct evaluation of whether reporting thresholds had been met or exceeded (e.g. pounds of chemicals used annually).
- **Threshold Estimates/Secondary Information** - secondary information such as sales, production volume, number of employees, etc. was collected in order to investigate correlations and relationships between it and reporting thresholds. Also, existing models (e.g., Regulatory Impact Analyses) which would enable estimation of threshold values were sought out.
- **Indicators** - descriptive statistics (average, median, maximum, minimum) were generated for secondary information collected on known reporters, comparison of these statistics to corresponding values for potential non-reporters may give an indication of whether a facility should be reporting.

The only primary information readily available was number of employees, which was available in Dun & Bradstreet and other commercial databases. The accuracy of employee size information in Dun & Bradstreet should be good since this information is self-reported by companies. Mitigating the reliability of this somewhat is the lack of employee size data for some facilities, employee size data which is only attributed to the company and not to individual facilities, and reporting errors (e.g. attribution of employees to a facility instead of the whole company).

Primary information on chemical usage was not available for any of the sectors. Chemical usage for facilities in the wood preserving sector was estimated using the models created for a 1990 Office of Solid Waste Regulatory Impact Analysis (RIA) done for the listing of wood preserving wastes.<sup>8</sup> The RIA contained models which broke down wood preservers by region, chemical preservative type, and production capacity. The AWPI survey of the wood preserving industry provided information on preservative type and production capacity of wood preserving facilities; this enabled selection of the

---

<sup>7</sup> Facilities that conduct manufacturing operations within Standard Industrial Classification (SIC) codes 2000 through 3999 are required to submit a TRI Form R if they meet the following thresholds: have 10 or more full-time equivalent employees, and manufactures (defined to include importing) or processes more than 25,000 pounds; or otherwise uses more than 10,000 pounds of any TRI-listed chemical during the calendar year.

<sup>8</sup>U.S. EPA, Office of Solid Waste. *Regulatory Impact Analysis for the Final Listing of Certain Wood Preserving Wastes* ("RIA"), prepared by ICF Incorporated for the Economic Analysis Staff, Office of Solid Waste, U.S. EPA, November 1990.

appropriate model for each facility. Coupled with sales information from commercial databases it was possible to generate an estimate of a facility's chemical usage. Regulatory impact analyses, or other comprehensive sector models, were not available for the other sectors.

For the data set which included facilities in a sector that had reported to TRI during a given year, it was hypothesized that there should be a relationship between annual chemical use and secondary information data elements. It was further hypothesized that the strongest link should be between chemical usage and production volume, followed by sales, and number of employees. If a strong relationship existed between annual chemical usage and one of the secondary information data elements for the reporters, then this relationship could be used to estimate potential non-reporter chemical usage for which only the secondary information data element was available. Linear regression analysis was conducted for the first two sectors, wood preserving and paper/paperboard production. For the wood preserving sector, the estimate of annual chemical usage derived from the RIA was used in the analysis. For the paper/paperboard sector, the quantity of chemicals reported in annual TRI submissions was used.

For the wood preserving sector, known TRI reporters were broken into three categories corresponding to major chemical process types in the industry (creosote, pentachlorophenol, and waterborne preservatives). Analysis was done to determine whether there was a strong relationship between chemical usage and the variables sales and number of employees. A strong linear relationship was found between the annual Creosote usage and number of employees. However, the subset of wood preserving facilities that used Creosote was relatively small, and there were few gaps in the chemical usage data for the model to fill. No other strong relationships were found for the wood preserving or the paper/paperboard sectors. One explanation for not finding the expected relationship may be the lack of specificity of the model (e.g. type chemical preservative used may not define a homogeneous set of producers for which a relationship exists).

When it was not possible to estimate chemical usage via an existing model or a linear relationship, several simpler statistical methods were employed as indicators of reporting status. Descriptive statistics (average, median, maximum, minimum) were calculated for sales, production volume, number of employees, etc., broken out by known TRI reporters and potential non-reporters. Comparison of values for non-reporters to the average and median values for reporters gave an indication as to whether or not a facility that did not report in a given year should have. For example, if the average sales for a reporting facility was \$150,000, and a non-reporting facility had sales of \$500,000 this may be an indication that the facility should have reported. For the paper/paperboard sector an analysis of PCS, AFS, and RCRA emission/effluent levels by reporters and non-reporters was done to determine if emission/effluent levels reported to other media programs could be used as an indicator of whether facilities should be reporting to TRI. However, a strong correlation between TRI reporting status and emission/effluent levels reported to other media programs was not found, and this technique was not used. Whether a facility had reported to TRI in a year prior to the one under investigation was also

used as an indication that a facility may be a potential non-reporter.<sup>9</sup>

Once the analysis was complete data sheets were generated for each facility in a sector which presented background information on the facility, reporting status, results of analysis using each of the methodologies, and summary statistics for the sector. Facility-specific data sheets were designed to present known information on a facility in one location in order to facilitate decision-making and further analysis. Samples of facility specific data sheets are shown on the following two pages. Additionally, on a regional basis, facilities were ranked in terms of likelihood of being a non-reporter. The ranking was based on the number of indicators which identified the facility as a potential non-reporter, the greater the number of positive indicators the greater the likelihood that a facility was a non-reporter. See Appendix B for a sample of the facility rankings from each of the sectors.

---

<sup>9</sup> See Appendices E and F, Abt Associates, Inc. memorandums concerning the screening of the wood preserving and the paper/paperboard sectors respectively, for a more complete discussion of the data analysis techniques employed.

## Wood Preservers Facility Level Analysis

**Facility Name:** ABC Wood Preserver  
**Address:** 321 Wood Preserver Street  
**City:** Big Woods  
**State:** MT  
**Zip:** 67890  
**EPA Region:** 8  
**Facility Source\*:** LP  
**TRI Identifier:** 67890ABCWO321WO  
**FINDS Identifier:** BNM598346722

### Above Threshold?

Employees Yes

Chemical Use Yes

Has this facility filed a report in previous years?

Not available

### Facility Data

		Units	Source*
Preservative Type	Waterborne	—	AWPI, 1995
Chemical Quantity Used	229,515 waterborne**	pounds/year	Estimated
	109,020 chromium		
	42,460 copper		
	78,035 arsenic		
Employees at Facility	40	employees	DMI
Facility Sales	1,00,0000	dollars/year	Estimated
Volume of Wood Treated	695,500	cubic feet/year	Estimated
1995 TRI Reporter	Yes	—	

\*Sources of information and methods used to estimate values are presented in Definitions and Data Sources.

\*\*Assuming the predominant waterborne preservative is chromated copper arsenate (CCA), a chemical breakdown is provided. Waterborne may be something other than CCA. Analysis of reporters indicates that the quantity of chemical used for preservative W (and its constituents) may be underestimated by the model by an average of 21%. See Section II. Statistical Analysis Methodology.

### TRI Reporter Statistics

Comparison to: TRI reporters using Waterborne preservative

	employees	sales (thous. \$)	chemical quantity (thous. lbs.)			
			W	Cr	Cu	As
<b>THIS FACILITY</b>	40	—	230	109	42	78
average	63	14,592	96	45	18	33
median	30	7,473	76	36	14	26
maximum	1,600	100,000	569	270	105	194
minimum	2	187	17	8	3	6

## Paper Mill Facility Level Analysis

**Facility Name:** ABC Paper Mill  
**Address:** 123 Paper Mill Road  
**City:** Wood Chip  
**State:** MN  
**Zip:** 54321  
**Region:** 5  
**Facility Source:** LP  
**TRI Identifier:** 54321ABCPA123PA  
**FINDS Identifier:** ERT386450938

Above Employee Threshold?	Yes
Has this facility filed a report in previous years?	Yes

Facility Data		Units	Source
Employees at Facility	120	employees	DMI
Facility Sales	—	dollars/year	NA
Production	240	tons/day	LP
1996 TRI Reporter	No	NA	TRI
Primary SIC Code	2631	Paperboard mills	DMI

### 1996 TRI Reportable Releases by Program (PCS, CAA, TRI, RCRA)

Release Type	Total (lbs.)	# of chemicals >10,000 lbs.	# of chemicals >25,000 lbs.	Top Chemical (name)	Top Chemical (lbs.)
Water	—	0	0	—	—
Air Emissions	—	0	0	—	—
TRI Releases	—	0	0	—	—
Haz. Waste Gen.	—	NA	NA	NA	NA

### TRI Reporter Statistics

	employees	facility sales (millions \$)	daily production (tons)
<b>THIS FACILITY</b>	120	—	240
average	521	273	243
median	273	62	160
maximum	16,002	7,259	860
minimum	2	1	4

September 30, 1998

## PROJECT EVALUATION

This section of the report will describe the actions taken to evaluate the effectiveness of the pilot project methodology at identifying TRI non-reporters in the wood preserving sector and the paper and paperboard sector. The evaluation plan consisted of conducting phone interviews and on-site inspections in order to test the accuracy of facility information and the success of the pilot methodology at identifying TRI non-reporters. Region 4 agreed to review the results of the wood preserving sector analysis and Region 1 agreed to review the results of the paper and paperboard sector analysis.

**Region 4 Review of the Wood Preserving Sector:** OPPA provided Region 4 with a spreadsheet with data collected on wood preserving facilities in Region 4 sorted by state, by inspection priority in the region, and by inspection priority in the state of Georgia. Due to regional resource considerations the inspections conducted in Region 4 were limited to the state of Georgia. After the region conducted their own review of the list of potential non-reporters seven Georgia facilities were targeted for inspection. Of the seven facilities inspected none were found to be non-reporters.

Of the seven facilities chosen for inspection three were identified as wood preservers in AWPI's 1995 Wood Preserving Industry Production Statistical Report, three were from Dun & Bradstreet Market Identifiers (DMI) database, and one was from EPA's facility indexing system. Inspections revealed that of the three facilities from DMI, two were not wood preserving facilities, while the third was a wood preserving facility, but had only 4 employees. It should be noted that for the DMI facility that was correctly characterized as a wood preserver, the pilot project was unable to find any information on number of employees, chemical usage, or annual sales. Therefore, its selection as an inspection target was questionable in light of the existence of other potentially non-reporting facilities for which more complete information was available, that gave a stronger indication of being an actual non-reporter. All three of the facilities identified from the AWPI report were correctly characterized as wood preservers. Of the three, one only had 8 employees, and one was under the chemical usage threshold.<sup>10</sup> The third facility had filed a report in 1997, however, the pilot project was attempting to identify 1995 non-reporters, so this finding is inconclusive. The one facility identified from EPA's facility indexing system was an administrative office, not a production facility. It should also be noted that the pilot project was unable to find any threshold or indicator information for this facility, again making its selection as an inspection target questionable. These findings lend support to concerns about the quality of information in commercial databases, that SIC codes are often incorrectly assigned, that it is difficult to distinguish between types of facilities (e.g. administrative and production) in existing commercial databases, and it raises questions about the accuracy of employee size data.

---

<sup>10</sup>There is an interesting finding concerning the facility that was found to be under the chemical reporting threshold. The company has both a lumber yard and a wood preserving operation in the same location. They were informed by their wood preservative supplier that if they spun off the wood preserving operation into a separate corporation, with only one employee, then they could avoid EPCRA § 313 reporting requirements. The region investigated this facility and found that they had not exceeded the chemical usage threshold. The two separate corporations were consolidated again in 1999.

Though none of the inspected facilities turned out to be non-reporters, it should be noted that of the Georgia facilities selected for inspection only one appeared in the top 10 of the list of facilities in the region sorted by inspection priority, and only 3 were in the top 25. Thus, it would be difficult to draw an overall conclusion regarding the success of the pilot project in terms of identifying potential TRI non-reporters in the wood preserving sector. However, if further investigation of potential non-reporters were to be conducted based on the information gathered for the pilot project it would be recommended that the origin of facility information be taken into consideration when prioritizing facilities, with facilities identified by D&B given a lower priority than those from other sources.

**Region 1 Review of the Paper & Paperboard Sector:** The pilot project identified 80 facilities as being in the paper/paperboard sector within Region I, of these 23 were identified as reporters, leaving 57 potential non-reporters. Region 1 reviewed all of the facilities identified as potential non-reporters either through review of existing regional records, call screening, or inspections.<sup>11</sup> Review of regional databases identified one facility as a reporter, two were inspected in 1994 and found to have no reporting obligation, three had been call-screened previously and found to be below reporting thresholds, nine were found to be out of business, and no information could be found on another. Of the remaining 41 facilities, two were selected for inspection and the rest were call-screened. Based on results of the call screening process an additional five facilities were selected for inspection. Of the seven facilities inspected, one was found to be in violation for failing to report the use of glycol ethers in 1995, 1996, and 1997; the region is also continuing to look at two other facilities which may have been required to report anhydrous ammonia.

Inspection of facilities that had been previously call screened confirmed the compliance determination made during the call screening process. Inspections found that most facilities were aware of EPCRA 313 reporting requirements. Though it was also found that facilities had made errors when making threshold determinations, such as failure to recognize that a substance contains a listed chemical, and inventory miscalculations. Inspections also identified potential problems with other sections of EPCRA.

Those facilities that had reported in the past were no longer doing so because chemicals had since been de-listed and were not required to be reported, or the scope of reporting required for a particular chemical had been narrowed. Additionally, some facilities were no longer reporting because process changes had resulted in the use of non-listed chemicals. See Appendix C for the Region I report on the review of the paper/paperboard sector.

Facility sales and production volume were used as indicators of reporting status for the paper/paperboard sector. Region I review of paper/paperboard facilities showed no correlation between statistical indicators and TRI reporting status. The diversity of the industry in terms of products produced and chemicals used may be one of the primary factors that resulted in no correlation between the indicators and reporting status. Also, the inability to determine whether a facility was using

---

<sup>11</sup>See Appendix C for Region I's write up of their review of the paper/paperboard sector.

a listed chemical before the screening or inspection led to numerous facilities being contacted which had no reporting obligation. For example, of the 41 facilities screened or inspected, 22 were using non-listed chemicals.

Another difficulty encountered with the paper/paperboard sector screening brings into question the timeliness of the data available in the Lockwood-Post's Directory of the Pulp, Paper and Allied Trades. Data was acquired on 66 facilities from the directory. Of these, nine were no longer in business, and no information could be found on a tenth. Three of the facilities had been out of business from two to eight years, ceasing operations in 1988, 1991, and 1994. This means a little over 15% of the facilities identified in Lockwood-Post were out of business. Though this sample is not large enough to draw an overall conclusion about the quality of information in the Lockwood-Post Directory, it does point out the importance of corroborating information when possible. It should be pointed out that despite problems with timeliness of data the Lockwood-Post Directory did a good job of identifying facilities in the sectors of interest as compared to Dun & Bradstreet database. If other regions were to make use of the data to target TRI non-reporters in the paper/paperboard sector it is advisable to check the accuracy of the data by using a call screening process, similar to that employed by Region I, to ensure that facilities were using listed chemicals and were over the employee threshold.



## LESSONS LEARNED

This section of the report collects and expands upon the lessons learned that have been touched upon in earlier sections of the report, and will provide additional insights into the problems encountered while conducting the electronic enforcement pilot project. The purpose is to provide a complete accounting of the issues involved in conducting this type of screening, and to inform future efforts that are similar in nature.<sup>12</sup>

**Choosing a Sector to Target is not Straightforward:** choosing an appropriate sector for electronically-based screening and targeting can be difficult and resource intensive. Considerable amounts of time and money were spent narrowing down which sectors would lend themselves to this type of analysis. It is not advisable to choose a sector without first gathering general background information on the sector, estimating the magnitude of the potential non-reporting problem, characterizing the sector in terms of chemical usage, and determining the availability of quality information sources.

General background information on a sector should include an estimate of the size of the sector (number of facilities), the average size of facilities (number of employees), and, if available, aggregate statistics such as production and sales. The number of facilities in the sector will allow you to determine whether the sector is of a reasonable size in terms of the resources available to conduct the analysis, yet large enough to allow meaningful statistical analysis to be done. The average size of facilities in the sector will give an indication of whether most facilities exceed the reporting threshold for number of employees. Because of the small business exemption for TRI reporting, if the sector is dominated by smaller facilities then it may not be as attractive a candidate for electronic screening. The size of the sector is also important in estimating whether a potential non-reporting problem exists. For example, if the best estimate of the size of the sector is 200 facilities and the EPA databases shows 170 are reporters, then the sector is not a good candidate for analysis since the size of the potential problem (worst case) is relatively small. Depending on the geographic distribution it may be more appropriate to investigate the potential non-reporters in a more ad hoc fashion at the regional level.

Characterizing the sector in terms of chemical usage is also important. It should be determined what the types of chemicals are predominantly used in the industry sector under investigation, and whether they must be reported to the Toxic Release inventory. If possible, it is helpful to rank the top 5-10 most frequently used chemicals. An attempt should also be made to determine where, on average, facilities annual chemical usage is in relation to the reporting threshold. This can be estimated if aggregate usage information is available for the sector. Alternatively, the average annual releases by chemical type of facilities reporting to TRI can be calculated and used as an indicator. If the average annual chemical usage is below or near the reporting threshold then the sector may not be a good candidate for analysis.

---

<sup>12</sup> See Appendix D for Abt Associates, Inc. assessment of the electronic enforcement screening pilot methodology.

A survey of the information sources available on the proposed sector must be conducted in order to determine whether there is sufficient information available, which is of an acceptable quality and cost to successfully complete the enforcement screening process. In particular, information on chemical usage (either direct or estimated) was hard to find, and often was the primary consideration in deciding whether or not to choose a sector for the pilot project.

Finally, it is important to ensure that targeting a sector does not conflict with any existing EPA initiative either within OECA or outside of OECA. It is also important to consider the geographic distribution of facilities in the sector, at least in a rough sense early on, in order to begin working with the regions in order to coordinate the research and plan follow up to the screening effort. Overall, development of a reliably accurate sector profile prior to undertaking an electronic screening effort is necessary both for producing positive electronic screening results, as well as to avoid potentially high costs and delays to conduct this sort of analysis during screening.

**Not All Sectors As Defined by SIC Code Lend Themselves to This Type of Analysis:** In many cases, sectors described by four digit SIC codes are too diverse in terms types of chemicals used and production processes employed for them to be good candidates for an electronic enforcement screen. The pilot project was unsuccessful at finding direct information on facility chemical use for any of the sectors screened, and it is unlikely that this information is available for other sectors. Given this, whether a facility has exceeded the chemical use threshold must be estimated via statistical analysis or indicators. There is a much better chance of identifying a useful statistical relationship, or that indicators will be meaningful, for sectors that are homogeneous, or near homogeneous, with respect to chemical usage and production processes.

Sectors that are homogeneous with respect to chemical usage predominantly use one, or a small group of listed chemicals. If several major chemical types are employed then it should be easy to segregate the facilities in the sector based on chemical type. The ability to focus on one or a few chemicals keeps the scope of research manageable. In most cases there is no direct information available on chemical usage, so estimation, statistical analysis, or indicators must be used to make the chemical usage threshold determination, this is easier to do when the sector is homogeneous in terms of chemical usage. For example, the majority of facilities in the wood preserving sector could be segregated by the three major chemical types used, which allowed analysis of the sector to take into account these major divisions. Increased specificity in the modeling increases the probability that a relationship will be uncovered.

Homogeneity in production process and product make it much more likely that a strong and consistent relationship between chemical usage and other variables such as production or sales could be found. A strong relationship between chemical usage and other variables increases the likelihood that a statistical relationship can be derived which will enable estimation of chemical usage, or that a correlation between reporting to the Toxic Release Inventory and a chemical usage indicator such as production exists. The paper/paperboard sector was not divisible by major chemical types, and produces a wide range of final products requiring the analysis to be more general in nature, relying on indicators as opposed to an estimate of chemical usage.

An additional consideration is how listed chemicals are used by the sector being considered for study. Sectors that use the chemical in the production process, or integrate it into the final product, are better candidates for this type of analysis. The wood preserving sector is a good example, where the listed chemicals are incorporated into the final product. Thus, making it reasonable to hypothesize that there is a correlation between annual chemical use and production. On the other hand, sectors which use the listed chemicals in a secondary operation or incidental to the end product (e.g. maintenance, sanitation, refrigeration) are not as good a candidate. It is much less likely that there is a link between this type of chemical use and some other variable such as production. Also, the volume of chemical used is much more likely to vary across facilities depending on maintenance schedules, type and age of equipment, operating procedures, etc. The canned and frozen poultry is an example of the latter type of sector, where listed chemical usage was predominantly in sanitation and refrigeration activities. This type of usage makes it much harder to estimate chemical usage.

The only time that it would be advisable to select a heterogeneous sector for this type of enforcement screen would be if it was known ahead of time that a dependable source of information on chemical usage was available for the facilities in the sector.

**Universe Definition is Labor Intensive:** The pilot project found that the current state of data sources external to the EPA do not support a straightforward and relatively automatic compilation of the facilities in a sector.

Commercial databases hold out the promise of having the best coverage of facilities in a given sector. However, data quality and limited facility-specific information are currently impediments to a more aggressive use of this approach for identifying TRI non-reporters. Specifically, SIC codes for corporations and facilities that appear in commercial databases are assigned by the database providers, and are not sufficiently accurate to allow their use without some other method of confirming the SIC designation. Review of information for the wood preserving sector showed that facilities were often assigned the wrong SIC code. It was also found that many facilities were administrative or sales offices, not production facilities.

Databases which focus on a particular sector or group of industry sectors hold out the promise of higher quality information since the focus is narrower than the large commercial databases. These sources typically have information at the facility-level such as production statistics, and are much less likely to incorrectly identify a facility as being in the sector in question. However, review of information for the paper/paperboard sector, which relied on an industry-specific data source, identified a significant number of cases where facilities were out of business, some for a number of years; one facility could not be located at all by the region. This raises concerns with the timeliness and accuracy of the information in these types of databases.

The goal of comprehensively defining the universe for a specific industry sector was made more difficult because integration of information from EPA and outside data sources was largely a manual process. Once information from disparate sources was combined into a single database the process of removing duplicates and merging information was manual. Manual sorting and integration of data was

required due to inconsistencies in name and address information available from different sources. The process was made even more time-intensive since many of the industry-specific commercial databases, and almost all of the non-commercial industry specific data sources were only available in a paper format. This required significant manual data entry and proofing.

**Information in Commercial Databases is Does Not Support Facility-Level Analysis:** As explained in the previous section, general commercial databases do a poor job of assigning facilities the correct SIC code, and identifying production facilities. Further, these databases rarely contain facility-level information, and much of the information in the databases is estimated based on industry averages (i.e., may not be a reliable indicator of facility-specific sales or production volume). In general, commercial databases are geared towards corporate-level analysis, not the facility-level analysis required for this type of screening effort. On the other hand, industry-specific databases do a better job of identifying facilities in a sector, and most contain some amount of facility-level information.

**Other Federal Agencies Were Not Useful Information Sources:** Federal agencies collect and format information specific to their operational mission. It was found that extending the use of the information beyond its intended purpose is usually not possible. For the federal agencies contacted as part of the pilot project (see the chart on page 8) it was found that they had little information that was of use in conducting the pilot. The one exception was the census bureau, which supplied general information on the number of facilities in four-digit SIC sectors in the food industry. There is a good deal of overlap between the facilities regulated by EPA and OSHA, so OSHA was approached to see if they could assist in helping to define the universe of facilities for a particular sector. It was found that OSHA, like EPA, does not have a comprehensive list of facilities in a sector, but instead uses an ad-hoc targeting method. The Department of Agriculture has a comprehensive list of facilities in the meat and poultry sectors; however, the information lacks enough specificity to be able to assign a facility to a particular SIC code.

## RECOMMENDATIONS

The purpose of this project was to test the hypothesis that an electronic enforcement screening methodology would save time and resources by identifying potential non-reporters via electronic aggregation and analysis of existing data, which would enable limited inspection resources to be targeted more productively. Based on the current development of widely available external information sources, the experience conducting the pilot project, and the results from regional review of the potential TRI non-reporters, it is not feasible to blindly choose a sector and successfully employ this methodology to identify non-reporters. Electronic enforcement screening will be more successful when the target sector is homogeneous in terms of chemical use and production method, chemicals are incorporated into the end product, and good sources of facility-level information are available. Listed below are recommendations to improve future electronic enforcement screening efforts.

Future Projects Should be Opportunistic: When a good source of facility-specific chemical information is found for a sector, a strong statistical relationship between chemical use and some other variable, or a good indicator of reporting status, then this type of targeting analysis should be considered. Note, however, that being opportunistic in selecting a sector to be screened should not equate with being less than thorough in the analysis conducted.

Other EPA Analytical Efforts Should be Leveraged to Facilitate Electronic Screening Efforts: When planning or conducting analytical or data gathering projects consideration should be given to whether they can be used or extended to enable this type of screening effort. For example, Regulatory Impact Analyses (RIAs) typically develop models of the regulated industry that could be used to estimate chemical usage at individual facilities. The information gathered and models developed for the RIA could easily be extended with facility specific information in order to conduct an electronic enforcement screen.

Complete Definition of a Sector Universe Should not Be An Overriding Goal: In almost all cases, information sources currently available will not enable the universe of facilities in a sector to be quickly and easily defined. Instead, the focus of universe definition should be on a two or three quality information sources to identify facilities which are not currently in EPA databases. The difficulty encountered in completely defining the universe of facilities using currently available data sources does not bode well for other proposed efforts such as generating a statistically significant compliance rate, which hinges on an accurate definition of the universe in a sector.

Industry-Specific Data Sources Are Better Suited to Electronic Screening: For future efforts it is recommended that an industry specific data source (trade association or commercial) of acceptable quality be used as the basis for universe definition and other data needs, and that general commercial databases only be relied upon to fill information gaps. In an effort separate from a screening project, it may be worthwhile to catalog and evaluate external information sources, looking more closely at Federal and local government sources. This effort would benefit future Headquarters and Regional screening projects.

Chemical-based vs. Sector-based Initiatives: It may be possible to screen for reporting of the use of a particular chemical across industrial sectors as opposed to the chemicals used in a particular industrial sector. This will be possible when use of the listed chemical is fairly consistent across sectors, and there is a good source of information on facility specific usage of the chemical, or a relationship between use and some other variable such as production or sales. This approach has the additional advantage of focusing all the efforts on a particular chemical as opposed to a number of chemicals used in a given sector.

## **APPENDIX A**

### **INFORMATION SOURCES REVIEWED FOR THE PILOT PROJECT**

## Government Sources

### **U.S. Environmental Protection Agency**

#### ***Toxic Chemical Release Inventory System (TRIS) Database, accessed through EPA's Integrated Data for Enforcement Analysis (IDEA) System.***

TRIS is the database in which TRI reports and reporter information can be accessed. For additional information call TRIS User Support (703) 816-4434 or IDEA User Support (202) 564-2475.

*This source provided information on the TRI reporting status of facilities, and a list of facility names and addresses.*

#### ***Permit Compliance System (PCS), accessed through the EPA's IBM mainframe.***

PCS is an automated information management system maintained by the Office of Compliance to track permit compliance and enforcement status of facilities regulated by the National Pollutant Discharge Elimination System (NPDES) under the Clean Water Act. For additional information call PCS User Support (202) 564-7277.

*This a list of facility names and addresses, and an attempt was made to correlate information in the PCS system with TRI reporting.*

#### ***Automatic Information Retrieval System (AIRS) Facility Subsystem (AFS), accessed through the EPA's IBM mainframe.***

AFS is a subsystem of AIRS and contains emissions, compliance, and enforcement data on a facility-level. For additional information call AFS User Support 1-800-367-1044.

*This a list of facility names and addresses, and an attempt was made to correlate information in the AFS system with TRI reporting.*

#### ***Biennial Reporting System (BRS), accessed through the EPA's IBM mainframe.***

BRS maintains waste information for RCRA hazardous waste on a facility-level. Large quantity generators and facilities which treat, store and dispose hazardous waste are required to report to BRS every other year, 1995 being the most current data available. For additional information call BRS User Support 1-800-767-7274

*This a list of facility names and addresses, and an attempt was made to correlate information in the BRS system with TRI reporting.*

#### ***Integrated Data for Enforcement Analysis (IDEA) System***

IDEA is an EPA data management system developed in response to the need for integrated data on facilities involved in EPA enforcement and compliance actions; and is designed to retrieve data about facilities across program databases. For additional information contact



IDEA User Support at (202) 564-2475.

## **U.S. Department of Agriculture**

Meat and Poultry Inspection Directory, 1997

### Commercial Sources

#### **Dun & Bradstreet Marketing Services Database, accessed through EPA's Integrated Data for Enforcement Analysis (IDEA) System.**

Dun & Bradstreet Marketing Services Database is a commercial vendor product leased by EPA's Office of Information and Resources Management. For this project, a version called DMI (Dun's Market Identifiers) was used. This database is a comprehensive business directory which retains corporate and facility-level information. Information is compiled from various sources such as public records, wire services, banks, etc. The advantages to this external source are its electronic form and its ability to link to EPA databases using the D&B ID# and FINDS linkage tables. D&B includes sales volume and number of employees at the site.

Related operations such as warehouses and sales offices necessary to production within the selected sector are included. To address this, D&B has created a manufacturing flag which at present is poorly populated. SIC based analyses tend to return higher than expected number of facilities because of SIC code descriptions. The data is supposedly refreshed annually, but if a company has little activity or if no inquiries are made about the company, it may not be reviewed for several years. Contact Dun & Bradstreet Corporation (800) 424-2495 or IDEA User Support (202) 564-2475.

*This source provided facility names, employee counts, and company sales.*

#### **Manufacturers' News Inc. (MNI)**

Manufacturers' News Inc. is an industrial information source. They compile and sell electronic directories and databases for U.S. manufacturers that are available on diskette. MNI mails out surveys to every manufacturer it can find up to three times. If the third letter is ignored, the manufacturer is called directly. Records for each state are reviewed at least once a year. There is no linkage variable besides the company name and address. The main benefit to MNI is that it is available on disk. It also lists important fields such as annual sales, numbers of employees, and owner information. However, these fields are all included in Dun & Bradstreet. The product description field is very brief and basically derived from the SIC code. Directories are available from: Manufacturers' News, Inc., 1633 Central Street, Evanston, IL 60201-1569, phone (847) 864-7000, fax (847) 332-1100, e-mail [info@manufacturersnews.com](mailto:info@manufacturersnews.com), website [www.manufacturersnews.com](http://www.manufacturersnews.com).

*This source provided employee counts and sales (in ranges).*



***American Business Information Marketing, Inc. (ABI)***

American Business Information Marketing, Inc. collects data from a variety of sources, including Yellow Pages, Business White Pages, annual reports and 10-K filings for publicly traded companies. They assign SIC codes to facilities themselves, rather than rely on the facilities to make the assignments themselves. ABI procedures appear to be better than self reporting. The fields included in ASL's database do not provide any information on production capacity beyond sales, which is also in D&B. They report that records are reviewed at least every six months.

American Sales Leads (ASL) is a division of ABI. Earlier this year, ASL provided Abt Associates with a listing of facilities in the SIC code covering steel mills for Texas, which was compared to the known SFIP facilities in Texas. This experiment led us to conclude that ASL's SIC code definitions were inappropriate for our work. For example, a small welding shop doing some milling work was assigned the SIC code for steel mills.

ABI uses the same database that ASL uses. In addition, they do not sell their data on CD, but provide it on diskette on a per-record cost basis in response to subscriber queries. A query returning 500-100 records costs \$0.60 per record; a query returning 1,001-2,000 records costs \$0.48 per record. Also, they do not list a possible linkage variable such as D&B number. They did indicate that their proprietary ID variable, the ABI number, can be linked to a D&B number, at additional cost.

***Database Publishing Co.***

Database Publishing Co. offers an electronic database by SIC code of facilities with 20 or more employees, eliminating analyses of all smaller facilities. This source in electronic format is also considerably more expensive than the others: \$3,900. They include relatively detailed fields for products and product description, but no capacity or production indicator. There is no obvious linkage data such as a D&B number.

***Ward's Business Directory of U.S. Public & Private Companies, Gale Research Inc.***

A seven volume directory of businesses in the U.S., however the sample sent to Abt Associates was the record of a company with a Canadian address. A search for SIC code 2491 resulted in only 114 companies with no individual facility listings. The database contains company name, address, phone number, officers, and SIC code(s) and is only available in hard copy with no linkage mechanism. Companies typically contact Gale asking to be included in the directory. Gale occasionally contacts companies they hear of to ask them if they would like to be listed.

## Industry Specific Sources

### ***1996 Lockwood-Post's Directory of the Pulp, Paper and Allied Trades***

Lockwood-Post is a commercial database of Pulp, Paper and allied Trades updated yearly. Lockwood-Post lists facility-specific production volume, production equipment, products, steam/power generation, water usage, and effluent treatment. Contact Lockwood-Post at (415) 905-2200.

### ***The Directory of the Caning, Freezing, Preserving Industries, Edward E. Judge & Sons, Inc. - Publishers***

The Directory of the Caning, Freezing, Preserving Industries is an industrial information source. They compile and sell electronic directories and databases for U.S. manufacturers that are available on diskette. For additional information call Edward E. Judge & Sons at (800) 729-5517.

### ***American Wood Preservers Institute. (1995) The 1995 Wood Preserving Industry Production Statistical Report***

American Wood Preservers Institute is an industry organization that compiles statistics and publishes an annual survey of wood preservers. The AWPI Report and Directory lists 451 facilities involved in wood preserving across the U.S. One strength of this list of facilities is that it was compiled from several sources across which verification could be done. AWPI used member lists, information from the Department of Commerce, the US Labor Department, the EPA, and the *1996 Big Book* to compile the directory. The directory includes all commercial wood preserving plants found to be operational during all or part of 1995. AWPI feels that their list should be all-inclusive. The editors are involved in a continuous process of review in order to ensure that their data are accurate for their general industry statistical reports.

The directory is only available in a hard copy format. There are no numeric linkage mechanisms to FINDS or Dun & Bradstreet. Its most useful fields are the preservatives used and the total plant void (total volume of cylinders or retorts). They can be reached at: AWPI; 2750 Prosperity Ave; Fairfax, VA 22031; phone: (703) 204-0500

*This source provided a list of facility names and addresses, cylinder void volumes, and type of preservative(s) used.*

### ***U.S. EPA, Office of Solid Waste. Regulatory Impact Analysis for the Final Listing of Certain Wood Preserving Wastes ("RIA"), prepared by ICF Incorporated for the Economic Analysis Staff, Office of Solid Waste, U.S. EPA, November 1990.***

There were eighteen models presented in this analysis, which broke down the wood preserving industry by type of chemical preservative used, facility size, and geographic region.

*This source provided models which enabled estimates to be calculated of volume of wood*

*treated and quantity of chemicals used.*

### ***1997 Directory of the Wood Products Industry (DWPI), Miller Freeman Publishing***

The *DWPI* is intended to be a complete sales directory for the lumber and wood treating industry. The editors report contacting every plant in an effort to obtain the most complete data possible, however, their results are very spotty.

There are relatively few (264) facilities listed in this source. However, some facilities contain very detailed information as to the production and capacity of a plant. Therefore, it may be useful in the later stages of this project when investigating specific potential non-reporters. It is only available in paper and has no obvious linkage mechanism. In addition, some records have different fields than others. For example, one plant gives daily capacity while the other gives daily production. The facilities reporting both capacity and production indicate that they are not operating at full capacity, thus care must be taken if estimates of chemical use are to be made by production or capacity data.

### ***1997 Big Book, Random Lengths Publishing***

The *1997 Big Book* is another directory of the wood products industry. It is only available in hard copy. It is similar to the 1997 *DWPI* published by Miller Freeman but contains an estimated 90 additional entries for a total of about 350 facilities. The directory is intended to be a complete listing of facilities for sales purposes. It is typically in a facility's best interest to be included in such a directory. Random Lengths claims that their editors contact every facility in order to ensure a comprehensive listing. The facility records appear complete for the existing fields, however production/capacity and many facilities which appear on other lists are not included.

### ***1997 Members Directory, Southeastern Lumber Manufacturers Association***

The Southeastern Lumber Manufacturers Association (SLMA) publishes a directory of member companies' facilities. It contains a large section of lumber manufacturers plus a subsection of 51 treating plants in the southeast. The directory is intended to promote the products of SLMA members. The most important feature of this source is its production field. The only other source which lists actual production instead of capacity is *DWPI*, which is spotty. Although it has a small number of facilities listed, this directory might be useful to determine correlations between production, capacity, and chemical use. The data is complete for the fields listed, which focus on production, products, and services. The directory is only available in hard-copy, however since it is a small database this is less important.

***Sources of Supply of Pressure Treated Wood, Western Wood Preservers Institute***

Western Wood Preservers Institute (WWPI) publishes a list of their members as a promotional tool. It includes twenty-five wood preserving plants in the Northwest U.S. and is reviewed continuously. There is no production information other than that of types of treatment available.

***1997 Buyer's Guide, Southern Forest Products Association***

The Southern Forest Products Association (SFPA) publishes a list of members in the form of a buyer's guide. The guide includes a general section of wood products manufacturers with their associated services and a specific section of facilities involved in wood preserving. There are 30 facilities listed in the wood treaters section. The facilities are concentrated geographically in the southeastern United States, but there are member facilities in Arizona, North Dakota, and New York as well. The *Buyer's Guide* includes more specific information about products produced, but no measure of production, capacity, or amount of chemicals used.

**APPENDIX C**

**REGION I REPORT ON REVIEW OF PAPER AND PAPERBOARD SECTOR TARGETS**

**APPENDIX E**

**ABT MEMORANDUM, “ELECTRONIC ENFORCEMENT PROJECT: WOOD  
PRESERVING SECTOR FACILITY REPORTS PACKAGE”**